DOCUMENT RESUME

ED 175 113 EA 012 020

AUTHOR Taubman, Paul
TITLE Measuring Educational Benefits.
PUB DATE Apr 79
NOTE 26p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, California, April 8-12, 1979); Table 2 may not reproduce clearly

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Benefits; Elementary Secondary Education; Equal Education; Higher Education; *Income; Measurement Techniques; *Opportunities; Physical Health; Twins
IDENTIFIERS *Life Expectancy

ABSTRACT

The first study presented here looks at the effects of years of schooling on earnings and assesses how much earnings and years of schooling are attributable to inequality of opportunity. The second study examines the effects of schooling on health. The first study examined a sample of approximately 1,000 pairs of identical twins and 900 pairs of fraternal twins. Results suggested that education has only modest effects on earnings, with each year of schooling increasing annual earnings by less than 5 percent. Results also indicated that inequality of opportunity is more important as a source of inequality of schooling than a source of inequality of earnings. These results suggest that even if it were possible to eliminate all sources of inequality of opportunity, most of the inequality of earnings would remain. The second study used data from Social Security files concerning 90,000 individuals. Results indicated that those with a college degree have a greater life expectancy than less educated groups and that generally life expectancy rises substantially with education. Unlike previous studies, this study also found that these correlations applied even to subjects over 65. (Author/JM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P Taubman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED175113

EA 012 020

Measuring Educational Benefits

Paul Taubman*

It is often argued that education makes you healthy, wealthy and wise because education increases particular marketable skills, helps people make better informed and more intelligent decisions and to assimilate new information more quickly, and alters preferences towards various lifestyles. There are numerous empirical studies conducted by economists, medical doctors, psychologists, and sociologists whose findings are in conformity with this statement.[1] In this paper I will examine some new evidence on the healthy and wealthy parts of this trilogy. I will also examine the question of the extent to which variation in individual's earnings and schooling is attributable to inequality of opportunity.

To study the effects of education on earnings it is customary to compare the average earnings of individuals with different amounts of schooling. Of course for such a comparison to be valid, it is necessary that the individuals not differ on average in any other skills or attributes that affect earnings. Most people, expect that on average more highly educated people are more intelligent irrespective of schooling and that intelligence is also rewarded in the market place. Thus the estimates of the returns to schooling will be biased if cognitive abilities are not controlled. There may be other types of ability that are correlated with schooling and earnings which must must be controlled to obtain unbiased estimates of the schooling coefficient. Most studies on the returns to education have tried to control for these skills and personal attributes in several different ways discussed below.[2] In this

*Professor of Economics, University of Pennsylvania and Research Associate of the National Bureau of Economic Research, Incorporated. The research in this paper was supported by NSF Grant numbers SOC76-17673 and NIA Grant number 1-R01AG00913-01. The sections on wage returns to education and partially drawn from joint research with J. Behrman, Z. Hrubec, R. Pollak and T. Wales. The section on health is partially drawn from joint research with S. Rosen.

paper I will control for differences in personal characteristics by using a sample of twins. With such data, one can control for characteristics that are attributable to both genetic endowments and family or common environment. I argue that twins provide somewhat better controls than the skill measures currently available. I also argue that one can use data on twins to estimate the importance of inequality of opportunity, suitable defined.

The relationship of health to education has been studied far less than the relationship of earnings to schooling. In this paper I will measure poor health by death and will demonstrate that more educated people have substantially lower age specific death rates. In the work I report on here little attention is paid to controlling for other characteristics and variables that may partially explain the negative relationship of age specific death rates to education. The method proposed, however, greatly expands our ability to examine this issue.

Education and Earnings

Most studies in economics and many of those in sociology posit a human capital model to examine the impact of education on earnings. The basic idea in this model is that a person's pay depends on his skill level. How skillful he is depends on both his innate ability and on certain "investments" he or his parents can make. These investments come in many forms including parental expenditures of money, time, and love, and the individual's time spent in school, vocational training and the like.

A standard equation economists use to study earnings is one in which the natural logarithm of earnings depends on the years of schooling and the amount of other skills.

$$(1)\quad \ln Y_j = a + bS_j + cK_j + u_j$$

where Y is the wage rate or in some case earnings,

S is years of schooling,

K is all skills except those attributable to schooling,

u is a random error,

a, b, c, are coefficients to be estimated.

The level of unobserved skills, K, depends upon innate ability or genetic endowments (G) and all investments in human capital except schooling (N) as in equation (2).

(2). $K_j = dG_j + eN_j$

Many people have used variants of equation 1 to estimate b, the effect of schooling on earnings. During the 1960's and 1970's many ways have been tried to control for unobserved skills while estimating b. The most common approaches have been to include either direct measures of an ability such as an IQ score or variables such as parental education and income that are part of or correlated with the other investments, N. The second method used to control for unobserved skills has been sophisticated statistical methodology, an example of which can be found in Griliches and Mason.

To put our new results in perspective, it is useful to summarize briefly the previous findings on how important it is to control for unobserved skills. A much more thorough summary is given in Behrman Taubman Wales and Hrubec (1979) Chapter 1, and in Solomon and Taubman (1973). A few specific studies are given in footnote 2. While measures of IQ, parental education and income, number of siblings and religion are generally found to be statistically significant and have the signs one would expect, a majority of previous studies conclude that the estimates of b is reduced by 15 percent or less when controlling for these variables. The majority of studies such as Griliches and Mason (1972) and some of the results in Sewell and Hauser (1975) suggest that controlling for unobserved skills is not very important. However there are a minority of studies which have found the estimate of b to be reduced greatly, e.g., by 25 to 40 percent, when controlling for these variables. These include Taubman (1975, 1976), Fagerlind (1975), and Olneck (1977). It is possible to distinguish the majority and minority studies in several ways.

First most of my work falls into the minority group. Second in the minority group the sample members have 7 or more years of labor force experience. As far as I am aware only the Griliches and Chamberlain reanalysis of the Gorseline finds it unimportant to control for unobserved skills when men have substantial labor force experience. This sample, however, is based in the late 1920's. On the other hand studies which find controlling for unobserved skills to be unimportant are based on samples in which the people have few years of work experience. It seems to me that a plausible reconcilliation of the two sets of results is that it is important to control for unobserved skills once labor force experience exceeds five years because employers test for and learn about unobserved skills by observing on the job performance. I might add that the few longitudinal samples that have the requisite data and span the appropriate time period support this view.

Finally let me note that those studies which find controlling for K is important, often imply that schooling does not have a large pay-off in terms of earnings or wage rates.

Why Twins

A possible problem with the studies that use IQ tests and measures of family background is that other types of skills and characteristics such as motivation and a strong back that may affect K and vary by education level have not been held constant. It is partly to surmount this problem that colleagues and I collected data on a large set of twins.

There are two types of twins, identical and fraternal. Identical or monozygotic (MZ) twins occur when an already fertilized egg happens to split. Both pieces of the split egg are genetically identical, barring mutations. Fraternal or dizygotic (DZ) twins occur when 2 eggs are released from the ovary and each is fertilized by a separate spern. From biology we know that each gene contains 2 members. Each egg and each sperm contains a randomly determined member of each gene. When a sperm fertilizes an egg, the member

of each gene contributed by each parent combine to make up the two members of each gene of each offspring. Since there are many genes, the random selection of a member of each gene for each egg and each sperm assures that fraternal twins are not genetically alike. For both types of twins, the siblings also share some common environment in the womb and in the family after birth.

Let us substitute equation (2), which indicates how unobserved skills (K) are produced, into (1), which indicates how schooling (S) and unobserved skills (K) affect the ln of earnings (ln Y), rank the sibs in a twin pair alphabetically and then calculate the differences in a pair in schooling, ln earnings etc. Denote within pair differences on any variable by Δ e.g., $\Delta \ln Y = \ln Y_1 - \ln Y_2$ where 1 and 2 refer to the first and second sibling in a family. We can then express our equation as (3) $\Delta \ln Y = b\Delta S + (cd)\Delta G + (ce)\Delta N + \Delta u$
For identical twins G = 0 for all pairs. For both types of twins, Δ N consists of noncommon environment.

In a sample of identical twins, the within pair equations will control for all differences in unobserved skills arising from innate ability and from common environment. Differences in noncommon environment, however, will be unobserved and are not controlled. As long as it is not necessary to control for noncommon environment, equation (3) will yield the (unbiased) estimate of b we have been searching for. Note that this method will not work for fraternal or DZ pairs differences in innate ability, ΔG, are not zero. Griliches (1978) has demonstrated that it is possible for the (expected value of the) estimate of b obtained from within pair equation for lnY for DZ pairs, equation 3 to differ more from the true value of b than the estimate obtained from the equation for individuals, equation 1.

There are several reasons why MZ within pair equations might not yield unbiased estimates of b. First ΔN may be correlated with ΔS. It is possible for example that a person's schooling and other investments jointly depend on

whom a person marries and at what date. Also parents may be more willing to invest more in one child than another because of childhood illnesses and accidents, preference for one sib, etc. Behrman, Pollak and Taubman (1979) have examined this issue and conclude that one can not reject the hypothesis that the ΔN that influences ΔY and ΔS are uncorrelated for the MZ twins.

Second it is well known that random measurement error in independent variables will cause estimated coefficients (obtained from ordinary least squares) to be biased towards zero. But as Griliches has emphasized, the effects of measurement error are much more pronounced when using within sibling data than when using data on individuals.[3] I will illustrate the magnitude of the problem below.

The NAS-NRC Twin Sample

The twin sample I use is described in detail in Behrman, Taubman, Wales and Hrubec (1979) and in Taubman (1976). Briefly in 1955 the National Academy of Science - National Research Council (NAS-NRC) started a project to construct a twin panel as a resource in studying a variety of questions. The NAS-NRC initially collected birth certificates on nearly all white male twins born between 1917-1927. The twin panel consists of those pairs both of whose members are veterans. Our sample consists of those pairs in the panel who answered a 1974 survey. Even this brief description should indicate that the sample need not be a random draw from the population of twins. In addition there are a variety of reasons why twins need not be a random draw of the total population. For example twins are not reared in a single sib household.

Compared to white males of the same age cohort or to the veteran subsample, the respondents in our sample have above average earnings. However, our earnings equations yield very similar coefficients to those obtained from random samples of the population drawn in the same time period. We

are working on the issue of how generalizable our results are to the population of white males. In the meantime we can only hope that they apply to the larger group.

The sample consists of roughly 1000 pairs of identical twins and 900 pairs of fraternal twins. For those pairs where both served in the Navy as enlisted men, the person's score on the General Classification Test is also available.

For both the within pair analysis and the analysis of variance to be described below, it is necessary to distinguish identical and fraternal twin types. In principle this can be accomplished by using a large enough battery of biochemical tests since a pair of identical twins must have the same blood type, RH factor, etc. on all such tests. Such tests have been done for a minority of the pairs. For the most part, however, the twins have been classified on the basis of their answers to the following question: "In childhood were you as alike as peas in a pod or only of ordinary family resemblence?". This may seem like a very delicate item on which to base our analysis, Jablon et. al. (1967) have shown that for this sample peas in a pod are 95 percent as accurate as biochemical tests. I might add that I was brought up on canned peas, and I don't know how alike peas in a pod are. Indeed I have been told that there is a minor crisis in twin research since many more children and younger adults are as gastronomically deprived as I was and currently respond incorrectly to this question.

The Effect of Schooling on Earnings

In a large number of studies in economics, the semilog specification given in equation 1, amended to include years of work experience, has been used. Table 1 contains a summary of some sample regression results calculated across individuals and within pairs for 1973 earnings when the men were about 50 years old. When as in line 1 we treat all the individuals in the sample as unrelated individuals and when we control for no other

Table 1

Coefficient on Years of Schooling in Equation for ln of 1973 Earnings

Equation Based on	Number of Observations	Coefficient on Yrs./Schooling	t statistic on coefficient	Other variables held constant
1. Individuals	3870	.080	32.4	none
2. Individuals	3870	.069	25.8	A, B
3. Navy Pairs	404	.051	6.0	A, B, C
4. Within DZ Pairs	914	.059	8.3	B
5. Within DZ Pairs	1022	.026	3.5	B

Other variables held constant, A is: age, number of sibs alive 1940, Father's Years of Schooling, Mother's Years of Schooling, Father's Occupational Status (Duncan Score); and the following variables which are coded as (0,1) dummies) raised in rural area, raised as a Catholic, raised as a Jew, born in the South.

B is married in 1974

C is score on Navy General Classification Test

variables, the coefficient on years of schooling is a highly significant .080, which is unchanged when age is added. This equation's estimated coefficients on years of schooling is similar to that obtained from nationwide random samples collected in 1960 or later. See for example Lillard and Willis (1978) or Jencks et. al. (1972). From line 2 we see that holding constant a whole host of observed family background variables and marital status reduces the coefficient by about 12%. For those pairs where both were in the navy, we have data on the General Classification Test which is primarily a vocabulary test and is a measure of cognitive skills. Controlling for this test and for the variables in the previous line, and estimating the equation across pair averages reduces the education coefficient to .051, a reduction from line 1 of about 35% and from line 2 of about 25%.

In line 4 the estimated coefficient from the within DZ pair equation is a highly significant .059. This is about 25% less than the estimate in line 1 and intermediate between line 2 in which background and line 3 in which background plus IQ are controlled. Line 5 presents the within MZ pairs results. Here the coefficient on schooling plummets (to a still statistically significant) .026 which is 1/3 of the estimate in line 1 and a half of the estimate in line 3. Unfortunately this .026 is almost surely biased towards zero by measurement error. Yet other evidence suggests the variance in measurement error in schooling is no greater than 10% of the total variance in schooling. Correction for measurement error of this magnitude or less would result in a coefficient for schooling of .045 or less. While .045 is nearly 100% greater than the estimate in line 5, it is also about 50% of line 1 estimate, once adjusted for the same degree of measurement error. Thus in studying the effects of schooling on earnings in this sample, it is crucial to control for genetic endownments and family environment. It also appears that measures of cognitive skill and certain aspects of family background provide fairly adequate controls in that the estimate in line 3 is .05 which is

quite close to .045. However the estimate of .05 in line 3 would also be biased towards zero by measurement error. Thus there must be other abilities omitted from equation 3 with the biases from these omitted variables and measurement error approximately offsetting one another. Based on these findings, it appears that identical twins provide the best controls for studying the returns to education. However a good measure of cognitive skills and a battery of background variables do nearly as well. Since it is probably easier and cheaper to design and collect random samples of individuals with the latter data than samples of twins, this is not a trivial conclusion.

Once we adjust for measurement error, our estimates of the size of the bias is of the same order of magnitude as that found in other studies in which men have substantial amounts of labor force experience. Of course if we were to adjust the other studies for measurement error, they would show somewhat less omitted variable bias, but this is just another way of saying that identical twins are better as controls.

The above results suggest that education has only modest effects on earnings since each year of schooling increases annual earnings by less than 5%. Indeed if increased wages were the only benefits of education, it is doubtful if education would be a wise investment. There are, however, other benefits of education, one of which we will discuss shortly.

Equality of Opportunity

In this section I will indicate how data on twins can be used to estimate the contribution of inequality of opportunity to inequality of outcome, as measured by the variance of schooling and of earnings. I will define inequality of opportunity to occur when different individual's must pay different prices for the same item. In general it is thought that children from wealthier familes can borrow at a lower interest rate than other children in the absence of government loan programs.

This type of definition of inequality of opportunity, while not accepted by all economists, is used by many including Becker (1975), Bowen (1977), and Okun (1975).[3] Economists tend to adopt this definition of inequality of opportunity rather than one that focuses on equality of test scores or earnings because of its implications within the general economic framework of examining income redistribution. This framework is one in which improvements in fairness in the distribution of income generally require losses in economic efficiency or in other words a lower value of Gross National Product, GNP. GNP is the value of all produce in the economy. However policies which allow children of poor families to borrow at the same interest rate as children of rich familes not only can improve fairness but can also improve efficiency and raise GNP. This is not to deny that other policies can compensate for poor innate ability and that the cost of some such policies may be no greater than the price of eyeglasses. But compensatory policies will improve fairness at the cost of decreased economic efficiency and GNP. Many compensatory policies are clearly worth this cost but many others are not. However since reducing inequality of opportunity, as defined above, involves no costs, Okun voices the hope that inequality of opportunity is a large source of inequality of outcome.

The model I have in mind is one in which investments in human capital occur primarily in childhood. Parents can be thought as having an investment in human capital function for each child. For the first child this can be expressed as a linear function of the price of investment, interest rate, family wealth, parental tastes, and the child's genetic endownments, and random events.

$$(4)\quad I = \alpha + \beta P + \gamma i + \delta W + \lambda T + \psi G + v$$

where I is by parent or child in health and all other items that increase earnings,

P is the price of this investment

i is the cost of borrowing funds,

W is the parent's income or wealth,

T is the parent's taste, i.e. how well they like the child

v is a random variable.

The same function can be written for the second twin though T and v could vary for the 2 children in a family. As stated previously innate ability is the same for sibs who are identical twins but differs for sibs who are fraternal twins.

We can, of course, calculate the cross twin correlation, $\sigma I_1 I_2 / \sigma I_1 \sigma I_2$, in schooling or in any other observed investment in human capital. Suppose the only variables on the right hand side of (4) were W and v. Then the cross twin correlation would be positive because W would be the same for both sibs but would differ across families. The correlation would be less than 1 since the v would be uncorrelated within and across families. The other variables in equation 4 would also affect the magnitude of the cross twin correlation. It seems likely that P and i would be the same within families and vary across families. T need not be perfectly correlated though there is some evidence that suggests that parents are equally concerned about each twin which implies that T is the same within a family. Finally for DZ twins C is not perfectly correlated.

The cross twin correlation coefficient measures the degree of similarity of twins. The size of this correlation will depend on the coefficients and the cross twin covariances of all of the unobserved variables on the right hand side of equation 4. The unobserved elements on the right hand side of (4) can be divided into genetic endownments, those variables that are (mostly) constant within but not between families which we call common environment, and the random variables, or noncommon environment, v.

Using this categorization, we can substitute the investment equation, equation (4), into the skill production function, equation (2), and with proper choice of scale express K for the first child as:

$$(5)\quad K_{1m} = G_{1m} + N^{C}_{m} + N^{T}_{1m}$$

It should be noted that in equation (5) the G term incorporates the direct effect of innate ability on the unobserved skills given in equation (2) and the indirect effect that flows through investment, I, in equation (4). This type of formulation varies from the standard model professed by most geneticists, who have not included the effect of innate ability on I as a genetic effect, but as far as I can determine my treatment is consistent with the mathematics of the problem as developed initially by R.A. Fisher.

Equation (5) has K as its dependent variable, but K is as unobserved as G and N. However the same type of logic can be used to derive a similar equation for S and ultimately for Y though different G's and N's may be involved for each observed variable. Using a variety of techniques and certain assumptions, it is possible to use the information on cross twin correlations calculated for individuals and across twins separately for the identical and fraternal twins to estimate the contribution of G, N^{C} and N^{T} to the variance of S, Y and other variables. The details of the method are spelled out in Behrman, Taubman, and Wales (1977) and in Behrman, Taubman, Wales, and Hrubec (1979). There are several key assumptions. Perhaps the most crucial is that the expected value of the ratio of the variance of common to the variance of total environment ($\sigma^{2}_{N}C/\sigma^{2}_{N}$) is the same for both twin types. If this assumption is replaced by one that says the ratio for MZ twins is higher by a factor of Q, it is always possible to find a Q such that the contribution of genetics falls to zero.[4] However the general reason for MZ twins to share a more common environment is that their parents and themselves partially base investments on each twin's genetic endowments.

Since this dependence is allowed for in the model, this criticism as advanced for example by Goldberger (1977) is not valid when studying inequality of opportunity.

The other major assumption in the model is that inequality of opportunity for all types of investments in human capital can be traced back primarily to a single source namely differences in family income adjusted for tastes. There is some evidence to support this assumption.

Using a nonlinear maximum likelihood program we have estimated various versions of the model given in Table 2. This model traces the path to labor market success from schooling, to status of initial occupation, to status of 1967 occupation, to earnings in 1973. The coefficient on the observed variables, e.g., q and r, obtained from this maximum likelihood method are very close to those obtained from the corresponding MZ within pair equations. The new information is the estimate of the coefficients of the G and N variables, each of whose unobserved variances we can set equal to 1.

Table 3 contains Behrman, Taubman, Wales, and Hrubec's (1979) estimates of the contribution of the variance of the various G, N and u terms to the contribution of schooling, the two occupational status variables and the ln of 1973 earnings. In these estimates we define N to be common environment and u to be noncommon environment.

We estimate for example that 36%, 41% and 23 of the variances in schooling is attributable to genetic endowments, common environment, and noncommon environment respectively.[5] Our special concern is with the σ_N^2 row. Common environment explains 41 and 12% of the variance in schooling and the ln of earnings. There are certain technical factors such as measurement error which may have caused us to understate the effect of N, but Taubman indicates that correction for these difficulties would have only modest effects, e.g., the 12% would be no more than 17%. These results indicate that inequality

Table 2

Latent Variable Model

Structural Equations	Reduced Form
1) $Y_1 = S = aG + bN + u$	$Y_1 = S = aG + bN + u$
2) $Y_2 = OC_i = c'G + d'N + eG_1 + fN_1 + gS + u_1$	$Y_2 = OC_i = cG + dN + eG_1 + gu + u_1$
3) $Y_3 = OC_{67} = h'G + j'N + k'G_1 + mG_2 + n'N_1 + pN_2 + qOC_i + rS + u_2$	$Y_3 = OC_{67} = hG + jN + kG_1 + nN_1 + pN_2 + qu_i + ru +$
4) $Y_4 = \ln Y_{73} = s'G + t'N + v'G_1 + w'G_2 + xG_3 + y'N_1 + z'N_2 + \alpha N_3 + \beta OC_{67} + \gamma S + u_3$	$Y_4 = \ln Y_{73} = sG + tN + vG + xG_3 + yN_1 + zN_2 + \alpha N_3 + \lambda u_1 + \beta u_2 + u_3$

where

$c = c' + ag$

$d = d' + bg$

$h = h' + ra + qc$

$j = j' + rb + qd$

$k = k' + qe$

$n = n' + qf$

$s = s' + \beta h + \gamma a$

$t = t' + \beta j + \gamma b$

$v = v' + \beta k$

$w = w' + \beta m$

$y = y' + \beta n + \gamma b$

$z = z' + \beta p$

$\delta = \gamma + \beta r$

$\lambda = \beta q$

S is years of schooling

OC_i is occupational status in initial civilian occupation

OC_{67} is occupational status in 1967 occupation

$\ln Y_{73}$ is log of earnings in 1973.

of opportunity is much more important a source of inequality of schooling than of earnings and that even if it were possible to eliminate all sources of inequality of opportunity, most of the inequality of earnings would remain. Again this does not mean that society can not or should not engage in compensatory training or transfer programs but that such programs entail decreases in economic efficiency. Finally let me add that my estimates in the previous section suggest that while a compensatory schooling policy can be used to reduce earnings inequality, schooling is not a very powerful tool, as is evidenced by the size of its coefficient in Table 1.

Education and Life Expectancy

The results in the previous two sections may well be discouraging to many people. Let me turn, therefore, to a new topic which may prove to be more pleasant. Sherwin Rosen and Taubman (1979) are currently conducting a study on the socioeconomic determinants of age specific death rates or (if that is too morbid) life expectancy. The model we have in mind is one in which a person can affect his health by expenditures on medical care and choice of occupation, residence, etc. In this model health enters directly into his utility function and affects his income opportunities. Education can be important in this model for a variety of reasons. First education does have some impact on earnings, as just indicated. Those with more income may well purchase more and better medical care. Second those with more education are on average brighter and have more knowledge. Thus the more educated may process information better and adopt new techniques and regimes more quickly. Third the more educated generally work at different jobs and may be exposed to fewer health hazzards, work related accidents, and stress. Fourth the more educated may prefer a life style which generates greater health as a byproduct. Fifth the more educated may value health more highly and purchase more health care than the less educated who have the same income and face the same price schedules.

Table 4

Sources of Variance of Schooling, Initial and Later Occupational Status and Earnings

Percent of Total Arising From	S	OC_i	OC_{67}	$\ln Y_{73}$
$\sigma^2_{G_1}$	36%	08%	11%	10%
$\sigma^2_{G_2}$		23	03	03
$\sigma^2_{G_3}$			15	*
$\sigma^2_{G_4}$				32
$\Sigma\sigma^2_{G_i}$	36	31	29	45
σ^2_N	41	22	13	12
$\sigma^2_{u_1}$	23	02	04	01
$\sigma^2_{u_2}$		46	01	*
$\sigma^2_{u_3}$			53	01
$\sigma^2_{u_4}$				42

Source: Behrman, Taubman, Wales and Hrubec p. , $\lambda = 0.34$, $\rho^* = \rho' = 1$.

* implies less than 0.5%

Totals may not add to 100% because of rounding.

All these reasons suggest that the more educated will have lower age specific death rates which I will show shortly is correct. It, of course, would be desirable to sort out why the more educated have longer life spans since the various reasons given above have different policy implications. At this time, we have made only modest progress on such separation, but we have ambitious plans for the future.

There are available a number of studies of the relationship of education and mortality (See Brenner for a recent survey). Those studies that examine all causes of death are mostly based on interstate or other geographical comparisons. Such studies are very useful; however, with geographic data it is very difficult to sort out statistically the separate influence of many variables which tend to be highly correlated.

To get at such issues it is necessary to have a sample of individuals with socioeconomic and mortality information. Since age specific death rates are small at most ages, less than 1% per year, the sample also must be large or gather mortality over several years. Kitigawa and Hauser (1973) assembled such a sample by matching 300,000 death certificates for the period May through August 1960 with 200,000,000 people enumerated in the 1960 Census of the population. This was, as you may imagine, a costly and time consuming process.

Rosen and I have adopted an alternative technique. The Social Security Administration learns of and records the date of most deaths in the process of paying and terminating benefits in its various programs. It is possible to match random samples to the Social Security files and obtain the data on death, past earnings, and benefits provided it is not possible for the user of the tape to learn the actual identity of any individual. The 1973 Current Population Survey (CPS) which is a random sample of the noninstitutionalized population has been so matched and is available as a public use tape.

The sample has about 90,000 individuals of whom some 3000 dies during the next three years. The sample is too small to study mortality of nonwhites. A comparison with Vital Statistics data indicates that SSA mortality information is currently less complete for women than men. For this reason we shall concentrate on men. The mortality information is also likely to be less complete for nonretired individuals then retirees (because most retirees receive checks until they die) and for single men than married men, or for men with no dependent children than men with dependent children (because surviving widows and dependent children are entitled to benefits).

We have calculated death rates in various age groups for different levels of education and have divided theses rates by the rate for those with 16 or more years of schooling in the same age group. The results are shown in Table 5. In each age group those with a college degree have the lowest death rate; thus, these relative death rates exceed 1. Generally the relative death rates decrease with education. Compared to college graduates, men with an elementary school education have a 40% and 70% greater death rate for those 55-64 and 65+ respectively. For females over 65 those with an elementary school education have a relative death rate 40% greater than college graduates. These differences are surely large. Our death rate relatives also are different in an important sense from the results obtained from the 1960 Census by Kitagawa and Hauser.

For white males 25 to 64 years old they and we obtain roughly comparable results with the age specific death rate of the least educated group exceeding that of college graduates by about 50%.[6] For those over 65, they find no difference by education level. We, however, find that in this age group the least educated have substantially higher death rates.

The findings for those over 65 is particularly surprising once it is recalled that since 1960 medicare and medicaid and food stamps were introduced, and the real value of social security payments were increased. We expected

Table 5

Relative Death Rates by Education Level
(College Graduates = 1.0)

	Education Level				
	1-8	9-11	11-12	13-15	16+
White Males					
55-64					
# Died	101	59	81	32	26
# People	1555	929	1542	509	564
Relative Death Rate	1.7	1.4	1.1	1.4	1.0
65+					
# Died	687	143	196	50	64
# People	2591	745	883	290	422
Relative Death Rate	1.4	1.3	1.5	1.1	1.0
White Females					
65+					
# Died	372	116	148	54	34
# People	2986	1035	1476	490	386
Relative Death Rate	1.4	1.3	1.1	1.3	1.0

Source: Rosen and Taubman p.

$$\text{Relative Death Rate} = \frac{\text{Death Rate for the given education level}}{\text{Death Rate for College graduates}}.$$

these innovations and changes to be of greatest value for the poor and least educated. Thus we would expect smaller differences by education level in 1973 than in 1960. Another surprising finding is the sharp difference in results by age level in Kitagawa and Hauser's (1973) study.

We have pondered over these results and have come up with a possible explanation though we hasten to add that currently we don't have data with which to test it. The annual death rates for those under 65 are relatively small. It is possible for many of these deaths to be related to a person's occupation. This link can occur because of accidents, because of exposure to chemicals and other dangerous elements, or because of stress. In a modern, industrialized economy the more educated generally have different jobs and may be less exposed to health hazards and perhaps stress. This type of explanation can explain the similarity in findings for those under 65.

If exposure to certain chemicals etc. has a long term cummulative effect, it is also possible to explain the shift in results for those over 64 on the grounds that in 1960 this group spent relatively little time in the modern industrialized economy. That is those over 64 in 1960 were born no later than 1895 and for the most part ended their labor force experience prior to 1955. Those over 64 in 1973 were born and worked 13 years later. During that time interval the structure of the American switched, becoming more industrial and less agricultural. Moreover education and occupation became more tighly tied. Thus those over 64 who had low education were much more likely to be exposed to industrial health hazzards in 1973 than in 1960. Let me caution you however, that much work remains to be done before this hypothesis can be considered as proven.

It is possible to try to control for more variables while estimating the impact of education on mortality. We have estimated a number of linear probability models. While there are several variables besides education which are significant, the inclusion of most of these variables has little impact

on the size or significance of the education coefficients. However inclusion of variables that seem to measure poor health does lower the education coefficients. While this finding probably means that people with low education are more likely to be in a state of poor health before dying, the nature of the health variables suggest that the higher death rates for the less educated are not attributable to accidents or other "instanteous" causes of death.

Conclusion

Perhaps the easiest way to conclude this piece is by several very short statements. Education can affect a person's life changes both in the sense that the more educated receive higher earnings and live longer. However based on the analysis of identical twins, much of the observed differences in earnings of the more educated are attributable to the greater innate ability of the more able. The analysis of the data on twins also suggest that inequality of opportunity explains much of the inequality in attainment of schooling but only a modest amount of the inequality in earnings. This latter finding suggests that significant income redistribution will require some loss in economic efficiency.

The more educated live longer. Moreover for some reason the relative advantage of this group is growing in the sense that it now exists for those over 65. Hopefully future research will establish why the differential and its trend are found.

BIBLIOGRAPHY

Becker, G., Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, Second Edition, NBER, N.Y., 1975.

Behrman, J., R. A. Pollak and P. Taubman, "Patterns of Parental Investment in their Progeny, "University of Pennsylvania, mimeo, 1979.

Behrman, J., P. Taubman and T. Wales, "Controlling for and Measuring the Effects of Genetics and Family Environment in Equations for Schooling and Labor Market Success," in Kinometrics: The Determinants of Socio-economic Success Within and Between Families, P. Taubman ed., 1977.

Behrman, J., P. Taubman, T. Wales, and Z. Hrubec, Socioeconomic Success: A Study of the Effects of Genetic Endownments, Family Environment and Schooling, North Holland, forthcoming (1979).

Brittain, J., The Inheritance of Inequality Brookings Washington, D.C. 1975.

Chamberlain, G., and Z. Griliches "Unobservables with a Variance Components Structure: Ability, Schooling, and the Economic Success of Brothers" International Economic Review 1975 pp. 422 - 449.

Fägerlind, I. Formal Education and Adult Earnings Almquist and Wicksell, Stockholm, 1975.

Feldman, Kanel, Newcomb T., The Impact of College on Students, Jossey-Bass 1969 San Francisco.

Fisher, R. A., "The Correlation Between Relatives on the Supposition of Mendelian Inheritance," Trans. Roy. Soc. Edinburgh, Vol. 42, pp. 321 - 341.

Goldberger, A. "Twin Methods: A Skeptical View" in Kinometrics

Griliches, Z., and W. Mason, "Education, Income and Ability," Journal of Political Economy, 1972.

Griliches, Z., "A Partial Survey of Sibling Models, "Harvard University, mimeo, 1978.

Jablon, S., et. al., "The NAS-NRC Twin Panel: Methods of Construction of the Panel, Zygosity Diagnosis and the Proposed Use, "American Journal of Human Genetics, Vol. 19, 1967, pp. 133-161.

Jencks, S., et. al., Inequality: A Reassessment of the Effect of the Family and Schooling in America, New York, Basic Books, 1972.

Kitagawa, E., and P. Hauser, Differential Mortality in the U.S., Harvard University Press, 1973.

Lillard, L. and R. Willis, "Dynamic Aspects of Earnings Mobility," Econometrica, September, 1978.

Okun, A., Equality and Efficiency: The Big Tradeoff, Brookings, Washington, D.C., 1975.

Olneck, M., "On the Use of Sibling Data" in Kinometrics Determinants of Socioeconomic Success Within and Between Families, ed. Taubman, North Holland, 1977.

Rosen, S., and P. Taubman, "Changes in the Impact of Education and Income on Mortality in the U.S., "University of Pennsylvania, mimeo, 1979.

Sewell, W. and R. Hauser, Education, Occupation and Earnings: Achievement in the Early Career, New York: Academic Press, 1975.

Solmon, L. and P. Taubman, Does College Matter Academic Press, 1973, New York

Taubman, P., Sources of Inequality of Earnings, North Holland, Amsterdam 1975.

Taubman, P., "The Determinants of Earnings: Genetics, Family and Other Environments: A Study of White Male Twins, American Economic Review, December, 1976.

FOOTNOTES

1. See Behrman, Taubman, Wales and Hrubec (1979) Chapter 1 or Solomon and Taubman for lengthy lists of citations.

2. Behrman et al provide many references. Some examples include the use of a few family background measures and IQ measures in Sewell and Hauser, the use of similar measures and some sophisticated methodology in Griliches and Mason, and an extensive list of family background, IQ and attitude measures in Taubman (1975).

3. See Griliches, (1978), The basic reason for this result is that differences in measurement error are a larger proportion of within pair differences in schooling than measurement error is of levels of schooling.

4. See Brittain ( ) for an exception.

5. See Behrman, Taubman, Wales and Hruebec (1979), Chapter 4.

6. These estimates are from a model in which we estimate the DZ cross twin correlation coefficient for the G terms as .34.

7. We and they use slightly different methods to adjust for age differences. Also the 1960 Census includes the permanently institutionalized while the CPS does not. However from the available evidence, these differences should not greatly affect the comparability of the results.